

Introduction to Reference and Anaphora in Iberian Languages

Iker Zulaica Hernández, Carlos Gelormini Lezama

Within the field of linguistics, *reference* is broadly defined as the human ability to refer to physical, textual, and mental entities by means of linguistic expressions, also known as *referring expressions*. The human ability to refer is an essential component of our cognitive system. We could hardly conceive both the outside and our inner world, interact with our interlocutors, share experiences, plans, ideas, narrate past events, or anticipate the future without the ability to refer and displace reference. Linguistic reference has been one of the central topics in the fields of semantics, pragmatics, and philosophy of language for decades. More recently, linguistic reference has become a hot topic in the field of cognitive science in general and, in particular, in psycholinguistics.

Natural languages have different mechanisms to refer. In anaphora, speakers use a linguistic expression (the anaphor) in order to refer to another expression in the linguistic/textual context (the antecedent). Anaphoric expressions can be of a varied nature (pro-forms, deictic words, tense, etc.) thus giving rise to a multiplicity of direct and derived phenomena prone to investigation. As a linguistic phenomenon, anaphora is a central element to interpret discourse, to attain textual cohesion, to create coherence, and to enable the dynamic construction of meaning. It is difficult to conceive the study of linguistic reference without taking anaphoric processes into consideration.

However, despite the fact that many of the factors that underlie anaphora resolution have been arguably claimed to be universal, most research has been carried out in English. This constitutes a serious limitation if we hope to gain a better understanding of the general principles that affect anaphoric processing. Studying anaphora in other languages -especially languages with a different anaphoric inventory- can provide us with crucial information to evaluate and discuss the generality of these principles. This special issue is intended to serve that purpose.

This issue, *Reference and Anaphora in Iberian Languages*, explores the psycholinguistics of reference, anaphora, and discourse interpretation in some of the languages spoken in the Iberian Peninsula. It consists of eleven high-quality articles by renowned scholars from different international universities and research institutions. Our issue is intentionally wide in scope, thus covering different languages (Spanish, Basque, Catalan, Catalan Sign Language), linguistic subfields (semantics, pragmatics, syntax, discourse analysis, language acquisition, computational linguistics), methodologies, and theoretical frameworks.

Sarah Blackwell, a leading scholar in the field of discourse studies, opens this issue with a fascinating exploration into antecedentless anaphoric pronouns and null subjects in spoken discourse. In her paper, Blackwell addresses important questions in connection with how interlocutors are able to felicitously use and interpret non-canonical pronouns with the help of underlying cognitive frames, and with the adequacy of cognitive status hierarchies in their interpretation. Based on her analysis of conversations, she claims that neighborhood antecedents and the interlocutors' underlying discourse expectations are crucial in the use and interpretation of non-coreferring, minimally salient, pronouns. Blackwell's deep understanding of anaphora as a pragmatic phenomenon and of inferential pragmatic processes provides yet another important piece of research on discourse reference and interpretation.

This is the author's manuscript of the article published in final edited form as:

Hernández, I. Z., & Lezama, C. G. (2017). Introduction to Reference and Anaphora in Iberian Languages. *Journal of Psycholinguistic Research*, 1–4. <https://doi.org/10.1007/s10936-017-9489-8>

Joan Busquets offers the first systematic study on the Catalan pro-form *fer-ho* in a paper where he analyzes the different syntactic environments licensing or banning the use of this pro-form and compares it with the use and function of related anaphoric phenomena such as pseudo-gapping, gapping, and stripping. As a deep anaphora process, Catalan *fer-ho* takes a discourse referent as its antecedent allowing pragmatic control (i.e. reference is made to some object in a model of the world constructed by the interpreter via pragmatic inference), hence a proper interpretation of *fer-ho* requires adopting a semantic, discourse-based perspective. Busquets attributes specific discourse properties to the anaphor such as its ability to refer to contextually salient antecedents or to update topics introduced previously in the discourse. Based on its syntactic characteristics, he argues that the denotation of a *fer-ho* anaphor antecedent should be characterized as an event type (an abstract object). He also introduces rhetorical relations into the picture in order to explain the referential properties of orphan material in the anaphor with regard to structural parallelism and information structure.

Ceberio, Aduriz, Díaz de Ilarraza, and García-Azkoaga contribute to this volume with a thorough computational study on coreferential annotation in Basque. Working on an annotated corpus of Basque, the authors present a 26,000-word corpus tagged for coreferential relations, where they include a variety of coreferential structures such as pro-forms, proper names, ellipsis, nominal, and adverbial anaphors, in a range of syntactic environments. This piece of research addresses very important issues in the field of computational linguistics and natural language processing laying the foundations for the development of an application for semi-automatic tagging of coreferential relations and, ultimately, a valid standard corpus for an end-to-end coreference resolution system.

Referential processes are key in different research areas of discourse analysis such as rhetoric and argumentation; the expression of subjectivity/empathy; and discourse construction, to name only a few. In their paper, De Cock and González-Arias analyze political blogs with an aim to unveil how blog authors and commentators establish reference to self and to the interlocutors and how they use these strategies to position themselves and others in the digital public sphere. Through a close analysis of deictic strategies and construction of singular/plural identity mechanisms, the authors draw interesting conclusions. For example, they observed coincidences in the use of deixis in spoken language debates, an observed decline in the expression of individual identity in favor of group identity by commentators, and very specific patterns when it comes to addressing other interlocutors.

Marcos García-Salido puts to test third person subject pronouns with the aim to unveil the principal motivations for their use in a null subject language like Spanish. To date, two main explanations have been offered, namely, speakers use third person overt subject pronouns for referential purposes (expressing the subject and/or disambiguation) or for informational purposes (emphasis, contrast, focus). García-Salido carries out a qualitative and quantitative corpus analysis of conversational discourse. Based on the frequencies he found, he argues that neither the referential nor the informational properties of subject pronouns are sufficient to account for their expression. However, the author points in the direction of structural priming (activation of a structure by repetition) as the key factor to explain the use of overt third person subject pronouns in Spanish.

Carlos Gelormini-Lezama contributes to this volume with a critical review of the Repeated Name Penalty (RNP) and the Overt Pronoun Penalty (OPP) in Spanish. He shows that all existing data on these phenomena can be integrated into a unified pragmatic framework. Research on the RNP effect has shown that sentences containing repeated names are harder to

process than sentences containing pronouns when the antecedent of the anaphoric expression is the subject of the previous sentence. Subsequent research showed that such processing delay is eliminated if the antecedent is the object of the previous sentence. This effect, which occurs in English, Mandarin Chinese, and Spanish, may be a universal phenomenon. However, the Spanish results showed an additional effect, dubbed the OPP, whereby sentences containing overt pronouns were harder to process than those containing null pronouns when the antecedent was the subject of the previous sentence. Like the RNP, this extra processing cost was eliminated when the antecedent was the object of the previous sentence. The author reviews these two processing delays and suggests that both the RNP and OPP can be regarded as different superficial phenomena reflecting the same underlying cause, namely, an imbalance between discourse function and processing cost.

Natural language tenses are considered anaphoric since they appear to share many characteristics with pronouns. The Spanish imperfect aspect, as in other Romance languages like French, has been characterized as anaphoric insofar as its interpretation crucially depends on an 'antecedent' time (a reference time in the discourse context), hence its non-autonomous character. Within this context, Manuel Leonetti investigates the temporal anaphoric behavior of the Spanish imperfect aspect and the temporal anaphora-pronominal anaphora analogy. He provides very compelling arguments in favor of revisiting the non-autonomy of the imperfect and suggests that we should rather focus our attention to lexical aspect and rhetorical relations as an alternative way to explain the acceptability patterns of the imperfect when no explicit temporal antecedent is present. Leonetti concludes that the Spanish imperfect is not anaphoric *per se* and suggests that its anaphoric value emerges as a pragmatic inferential effect.

In his paper, Cristóbal Lozano offers an experimental study on the development of anaphora resolution in Greek learners of L2 Spanish at three proficiency levels (intermediate, lower advanced, upper advanced). To that aim, Lozano devises three resolution contexts based on discourse constraints known to affect the use of overt/null subject pronouns in Spanish: topic continuity, contrastive focus, and emphasis. Although Greek and Spanish are both null-subject languages sharing similar discourse constraints affecting the syntactic alternation of their overt/null subjects, Lozano shows that these similarities are not necessarily a facilitating factor in second language acquisition. The results of his study show that learners' sensitivity to anaphora resolution pragmatic constraints increases with proficiency and that learners are sensitive to contrastive-focus constraints but they are more relaxed with topic-continuity constraints. Lozano accounts for these results by postulating some general pragmatic deficits related to language economy principles.

Laia Mayol and Gemma Barberà compare the different anaphoric strategies used in Catalan and Catalan Sign Language (LSC) in one of the first comparative studies dealing with reference-tracking devices in spoken and sign languages. Using a parallel corpus, the authors closely examine overt subject pronouns in Catalan and the main factors triggering their use in discourse (expression of contrast, topic change, focus) with the aim to unveil how such anaphoric contexts are rendered in a language that exploits the visual-spatial modality such as LSC. Their study reveals special anaphoric strategies in LSC that include: Fewer use of pronouns compared to spoken Catalan; use of determiners directed to space and classifiers to build anaphoric chains in signed discourse; different strategies to mark topic changes; and the use of the manual sign 'also' as the main strategy to mark focus. In addition to dealing with complex and understudied issues in the field of discourse reference and anaphora in sign languages, the authors suggest new

possible research avenues having to do with the role of rhetorical relations in the interpretation of pronouns in spoken and sign languages alike.

The volume closes with Zulaica-Hernández's work on complement anaphora. The author provides a first approximation to this phenomenon in Spanish and argues that reference to complement sets via pro-forms in Spanish conforms to the general pattern observed cross-linguistically by means of which complement anaphoric readings are only possible with monotone decreasing quantifiers. Furthermore, he claims that the different proportional interpretations of certain quantifiers may be the key to explaining why some quantified expressions allow for multiple anaphoric readings (refset, maxset and compset interpretations) while other expressions do not. Zulaica-Hernández also explores the role that coherence relations may have in facilitating complement anaphoric readings, thus opening a new path for subsequent theoretical and experimental studies on complement anaphora use and interpretation.

Collective issues on referential and anaphoric processes in the languages of the Iberian Peninsula are simply nonexistent. The main goal of this collection of articles is to fill this gap through a detailed overview of the most current research on anaphora, reference, their interfaces, and on the production and interpretation of spoken and written discourse. Importantly, as stated above, anglocentrism has dominated the field of psycholinguistics and has undermined our ability to understand how universal principles and language-specific preferences may interact. We are confident that this volume will make a valuable contribution to the field of reference processing in general because of the quality of the articles, the diversity of analyzed phenomena, the breadth of theoretical paradigms, and the variety of methodological perspectives.

Iker Zulaica-Hernández and Carlos Gelormini-Lezama
The editors